



Generation of scenarios from calibrated ensemble forecasts with a dual ensemble copula coupling approach

Ben Bouallègue, Zied; Heppelmann, Tobias; Theis, Susanne E.; Pinson, Pierre

Published in:
Monthly Weather Review

Link to article, DOI:
[10.1175/MWR-D-15-0403.1](https://doi.org/10.1175/MWR-D-15-0403.1)

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Ben Bouallègue, Z., Heppelmann, T., Theis, S. E., & Pinson, P. (2016). Generation of scenarios from calibrated ensemble forecasts with a dual ensemble copula coupling approach. *Monthly Weather Review*, 144, 4737-4750. <https://doi.org/10.1175/MWR-D-15-0403.1>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Generation of Scenarios from Calibrated Ensemble Forecasts with a Dual-Ensemble Copula-Coupling Approach

ZIED BEN BOUALLÈGUE

Deutscher Wetterdienst, Offenbach, Germany, and Meteorological Institute, University of Bonn, Bonn, Germany

TOBIAS HEPELMANN AND SUSANNE E. THEIS

Deutscher Wetterdienst, Offenbach, Germany

PIERRE PINSON

Technical University of Denmark, Kongens Lyngby, Denmark

(Manuscript received 18 November 2015, in final form 8 September 2016)

ABSTRACT

Probabilistic forecasts in the form of ensembles of scenarios are required for complex decision-making processes. Ensemble forecasting systems provide such products but the spatiotemporal structures of the forecast uncertainty is lost when statistical calibration of the ensemble forecasts is applied for each lead time and location independently. Nonparametric approaches allow the reconstruction of spatiotemporal joint probability distributions at a small computational cost. For example, the ensemble copula coupling (ECC) method rebuilds the multivariate aspect of the forecast from the original ensemble forecasts. Based on the assumption of error stationarity, parametric methods aim to fully describe the forecast dependence structures. In this study, the concept of ECC is combined with past data statistics in order to account for the autocorrelation of the forecast error. The new approach, called d-ECC, is applied to wind forecasts from the high-resolution Consortium for Small-Scale Modeling (COSMO) ensemble prediction system (EPS) run operationally at the German Weather Service (COSMO-DE-EPS). Scenarios generated by ECC and d-ECC are compared and assessed in the form of time series by means of multivariate verification tools and within a product-oriented framework. Verification results over a 3-month period show that the innovative method d-ECC performs as well as or even outperforms ECC in all investigated aspects.

1. Introduction

Uncertainty information is essential for an optimal use of a forecast (Krzysztofowicz 1983). Such information can be provided by an ensemble prediction system (EPS) that aims at describing the flow-dependent forecast uncertainty (Leutbecher and Palmer 2008). Several deterministic forecasts are run simultaneously, accounting for uncertainties in the description of the initial state, the model parameterization, and, for limited area models, the boundary conditions. Probabilistic products are derived from an ensemble, tailored to a specific user's needs. For example, wind forecasts in the form of quantiles at selected probability levels are of particular

interest for actors in the renewable energy sector (Pinson 2013).

However, probabilistic products generally suffer from a lack of reliability, the system showing biases and failing to fully represent the forecast uncertainty. Statistical techniques allow users to adjust the ensemble forecast, correcting for systematic inconsistencies (Gneiting et al. 2007). This step, known as calibration, is based on past data and usually focuses on a single or a few aspects of the ensemble forecast. For example, calibration of a wind forecast can be performed by univariate approaches (Bremnes 2004; Sloughter et al. 2010; Thorarinsdottir and Gneiting 2010) or bivariate methods, which account for correlation structures of the wind components (Pinson 2012; Schuhen et al. 2012). These calibration procedures provide reliable predictive probability distributions of wind speed or wind components for each forecast lead time and

Corresponding author address: Zied Ben Bouallègue, Deutscher Wetterdienst, Frankfurter Strasse 135, Offenbach 63067, Germany.
E-mail: zied.ben-bouallegue@dwd.de

location independently. Decision-making problems can however require information about the spatial and/or temporal structure of the forecast uncertainty. Examples of application in the renewable energy sector resemble the optimal operation of a wind-storage system in a market environment, the unit commitment over a control zone, or the optimal maintenance planning (Pinson et al. 2009). In other words, scenarios that describe spatiotemporal wind variability are relevant products for end users of wind forecasts.

The generation of scenarios from calibrated ensemble forecasts is a step that can be performed with the use of empirical copulas. The empirical copula approaches are nonparametric and, in comparison with parametric approaches (Keune et al. 2014; Feldmann et al. 2015), simple to implement and computationally cheap. Empirical copulas can be based on climatological records [Schaake shuffle (ScSh); Clark et al. (2004)] or on the original raw ensemble [ensemble copula coupling (ECC); Schefzik et al. (2013)]. ECC, which features the conservation of the ensemble member rank structure from the original ensemble to the calibrated one, has the advantage of being applicable to any location within the model domain without restriction related to the availability of observations. However, unrealistic scenarios can be generated by the ECC approach when the postprocessing indiscriminately increases the ensemble spread to a large extent. Nonrepresentative correlation structures in the raw ensemble are magnified after calibration, leading to unrealistic forecast variability. As a consequence, ECC can deteriorate the ensemble information content when applied to ensembles with relatively poor reliability, as suggested, for example, by the verification results in Flowerdew (2014).

In this paper, a new version of the ECC approach is proposed to overcome the generation of unrealistic scenarios. Focusing on time series, a temporal component is introduced into the ECC scheme accounting for the autocorrelation of the forecast error over consecutive forecast lead times. The assumption of forecast error stationarity, already adopted for the development of fully parametric approaches (Pinson et al. 2009; Schölzel and Hense 2011), is exploited in combination with the structure information of the original scenarios. The new approach based on these two sources of information, past data and ensemble structure, is called *dual*-ensemble copula coupling (d-ECC). Objective verification is performed in order to show the benefits of the proposed approach with regard to the standard ECC.

The manuscript is organized as follows. Section 2 describes the dataset used to illustrate the manuscript as well as the calibration method applied to derive the calibrated quantile forecasts from the raw ensemble.

Sections 3 and 4 introduce the empirical copula approaches for the generation of scenarios and discuss in particular the ECC and d-ECC methods. Section 5 describes the verification process for the scenario assessment. Section 6 presents the results obtained by means of multivariate scores and within a product-oriented verification framework.

2. Data

a. Ensemble forecasts and observations

COSMO-DE-EPS is the high-resolution Consortium for Small-Scale Modeling (COSMO) EPS run operationally at DWD. It consists of 20 COSMO-DE forecasts with variations in the initial conditions, the boundary conditions, and the model physics (Gebhardt et al. 2011; Peralta et al. 2012). COSMO-DE-EPS follows the multimodel ensemble approach, with four global models each driving five physically perturbed members. The ensemble configuration implies a clustering of the ensemble members as a function of the driving global model when large-scale structures dominate the forecast uncertainty.

The focus here is on wind forecasts at 100-m height above ground. The postprocessing methods are applied to forecasts of the 0000 UTC run with an hourly output interval and a forecast horizon of up to 21 h. The observation dataset comprises quality-controlled wind measurements from seven stations: Risoe, *FINO1*, *FINO2*, *FINO3*, Karlsruhe, Hamburg, and Lindenberg, as plotted in Fig. 1. The verification period covers a 3-month period: March–May 2013.

Figure 2a shows an example of a COSMO-DE-EPS wind forecast at hub height. The forecast is valid on day 2 (March 2013) at *FINO1* (see Fig. 1). The ensemble members are shown in gray while the corresponding observations are in black. In Fig. 2b, the raw ensemble forecast is interpreted in the form of quantiles.

Formally, a quantile q_τ at probability level τ (with $0 \leq \tau \leq 1$) is defined as

$$q_\tau := F^{-1}(\tau) = \inf\{y: F(y) \geq \tau\}, \quad (1)$$

where F is the cumulative probability distribution of the random variable $Y \in \mathbb{R}$:

$$F(y) = \mathbb{P}(Y \leq y). \quad (2)$$

In practice, at each forecast lead time, the member of rank n can be interpreted as a quantile forecast at probability level τ_n :

$$\tau_n = \frac{n}{N_e + 1}, \quad (3)$$

where N_e is the number of ensemble members.

In the example shown in Fig. 2, the raw ensemble is not able to capture the observation variability. Calibration aims to correct for this lack of reliability by adjusting the mean and enlarging the spread of the ensemble forecast.

b. Calibrated ensemble forecasts

Since COSMO-DE-EPS forecasts have been shown to suffer from statistical inconsistencies (Ben Bouallègue 2013, 2015), calibration has to be applied in order to provide reliable forecasts to the users. The method applied in this study is the bivariate non-homogeneous Gaussian regression (EMOS; Schuhen et al. 2012). The mean and variance of each wind component, as well as the correlation between the two components, characterize the predictive bivariate normal distribution. Corrections applied to the raw ensemble mean and variance are optimized by minimizing the continuous ranked probability score (CRPS; Matheson and Winkler 1976). The calibration coefficients are estimated for each station and each lead time separately (local version of EMOS), based on a training period being defined as a moving window of 45 days.

The final calibrated products considered here are N_e -equidistant forecasts of wind speed estimated for each location and each forecast lead time separately, where the N_e probability levels associated with the forecast quantiles follow Eq. (3). Calibrated quantile forecasts are shown in Fig. 2c. The spread of the ensemble is increased with respect to Fig. 2b and thus the observation variability is now captured by the forecast. From a statistical point of view the calibration method provides reliable ensemble marginal distributions and reliable quantile forecasts as checked by means of rank histograms and quantile reliability plots (not shown). The performance of the applied calibration technique

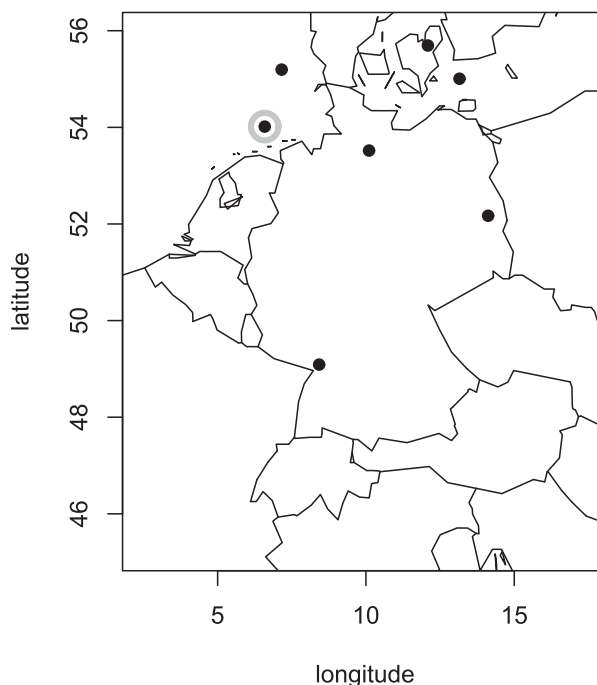


FIG. 1. Map of Germany and neighboring areas (approximately the COSMO-DE domain) with latitude–longitude along the axes. Locations of the seven wind stations used in this study (black circles). The station *FINOI* is highlighted with a gray circle.

is similar to the results obtained by other methods such as quantile regression (Koenker and Bassett 1978; Bremnes 2004).

Information about spatial and temporal dependence structures, which are crucial in many applications, are however no longer available after this calibration step (see Fig. 2c). The next postprocessing step consists then in the generation of consistent scenarios based on the calibrated samples.

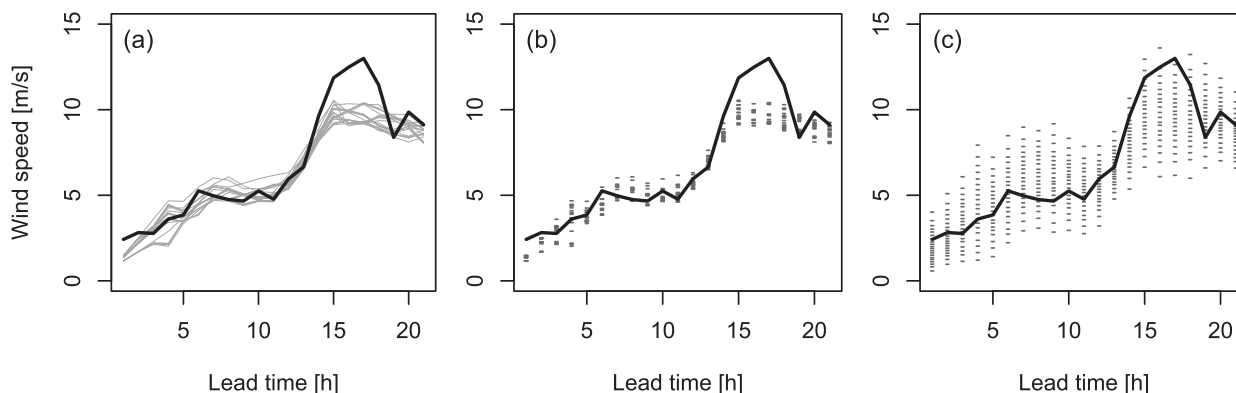


FIG. 2. Wind speed at 100-m height above ground (black solid lines) on 2 Mar 2013 at *FINOI* as a function of lead time (h): (a) COSMO-DE-EPS forecast (gray lines), (b) raw ensemble forecast in the form of quantiles (gray symbols, assorted members; see text), (c) calibrated quantile forecasts (gray symbols).

3. Generation of scenarios

The generation of scenarios with empirical copulas is here briefly described. For a deeper insight into the methods, the reader is invited to refer to the original article by Schefzik et al. (2013) or to Wilks (2015) and references within.

First, consider the multivariate cumulative distribution function (cdf) G defined as

$$G(y_1, \dots, y_L) = \mathbb{P}[Y_1 \leq y_1, \dots, Y_L \leq y_L] \quad (4)$$

of a random vector (Y_1, \dots, Y_L) with $y_1, \dots, y_L \in \mathbb{R}$. As in Eq. (2), we define the marginals F_l as

$$F_l(y_l) = \mathbb{P}[Y_l \leq y_l]. \quad (5)$$

Sklar's theorem (Sklar 1959) states that G can be expressed as

$$G(y_1, \dots, y_L) = C[F_1(y_1), F_L(y_L)], \quad (6)$$

where C is a copula that links an L -variate cumulative distribution function G to its univariate marginal cdf's: F_1, \dots, F_L .

In Eq. (6), a joint distribution is represented as univariate margins plus copulas. The problems of estimating univariate distributions and estimating dependence can therefore be treated separately. Univariate calibration marginal cdf's F_1, \dots, F_L are provided by the calibration step described in the previous section. The choice of the copula C depends on the application and on the size L of the multivariate problem. We focus here on empirical copulas since they are suitable for problems with high dimensionality.

An empirical copula is based on a multivariate dependence template, a specific discrete dataset \mathbf{z} defined in \mathbb{R}^L . The chosen dataset is described formally as

$$\mathbf{z} := \{(z_1^1, \dots, z_1^N), \dots, (z_L^1, \dots, z_L^N)\} \quad (7)$$

consisting of L -tuples of size N with entries in \mathbb{R} . In other words, L is the dimension of the multivariate variable and N is the number of scenarios. The rank of z_l^n for $n \in \{1, \dots, N\}$ and $l \in \{1, \dots, L\}$ is defined as

$$R_l^n := \sum_{i=1}^N \mathbb{I}(z_l^i \leq z_l^n), \quad (8)$$

where $\mathbb{I}(\cdot)$ denotes the indicator function taking a value of 1 if the condition in parentheses is true and 0 otherwise.

In practice, N equidistant quantiles of F_l with $l \in \{1, \dots, L\}$ are derived from the univariate calibration step:

$$\mathbf{q} := \{(q_1^1, \dots, q_1^N), \dots, (q_L^1, \dots, q_L^N)\} \quad (9)$$

with

$$q_l^n := F_l^{-1}(\tau_n); \quad n \in \{1, \dots, N\}, \quad (10)$$

where τ_n is defined in Eq. (3). The sample \mathbf{q} is rearranged following the dependence structure of the reference template \mathbf{z} . The permutations $\pi_l(n) := R_l^n$ for $n \in \{1, \dots, N\}$ are derived from the univariate ranks R_l^1, \dots, R_l^N for $l \in \{1, \dots, L\}$ and applied to the univariate calibrated sample \mathbf{q} . The postprocessed scenarios $\tilde{x}_l^1, \dots, \tilde{x}_l^N$ for each margin l is expressed as

$$\tilde{x}_l^1 := q_l^{\pi_l(1)}, \dots, \tilde{x}_l^N := q_l^{\pi_l(N)}. \quad (11)$$

The multivariate correlation structures are generated based on the rank correlation structures of a sample template \mathbf{z} . The empirical copulas presented here only differ in the way \mathbf{z} is defined. In the following, let $t \in \{1, \dots, T\}$ be a lead time and let $L \succ T$. For simplicity, we consider here a single weather variable and a single location.

a. Ensemble copula coupling

The rank structure of the ensemble is preserved after calibration when applying the standard ECC approach. The raw ensemble forecast is denoted \mathbf{x} :

$$\mathbf{x} := \{(x_1^1, \dots, x_1^{N_e}), \dots, (x_L^1, \dots, x_L^{N_e})\}, \quad (12)$$

where N_e is the ensemble size. ECC applies without restriction to any multivariate setting. The number of scenarios generated with ECC is however the same as the size of the original ensemble ($N = N_e$). The transfer of the rank structure from the raw ensemble forecast to the calibrated one consists then of taking \mathbf{x} as the required template in Eq. (7).

Based on COSMO-DE-EPS forecasts in Fig. 3a (identical to Fig. 2a), an example of scenarios derived with ECC is provided in Fig. 3b. The increase in spread after the calibration step implies a larger step-to-step variability in the time trajectories. Figure 4 focuses on a single scenario highlighting the difference between the original and postprocessed scenarios.

b. Dual-ensemble copula coupling

ECC assumes that the ensemble prediction system correctly describes the spatiotemporal dependence structures of the weather variable. This assumption is quite strong and cannot be valid in all cases. On the other hand, based on the assumption of error stationarity, parametric methods have been developed that

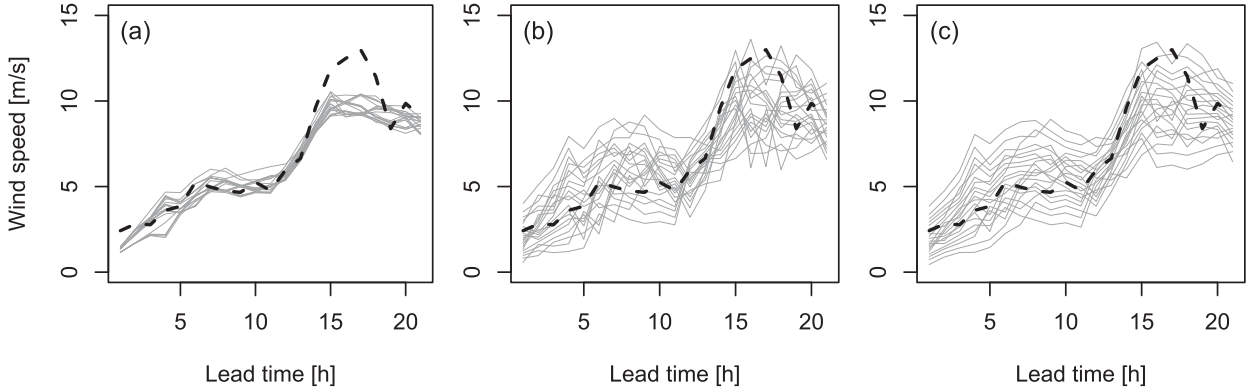


FIG. 3. As in Fig. 2, but for scenarios (a) COSMO-DE-EPS, (b) ECC-derived, and (c) d-ECC-derived, and the corresponding observations (black lines).

focus on the covariance structures of the forecast error (Pinson et al. 2009; Schölzel and Hense 2011). We propose a new version of the ECC approach, which is an attempt to combine both types of information: the structure of the original ensemble and the error autocorrelation estimated from past data. Therefore, the new scheme is called dual-ensemble copula coupling, as the copula relies on a dual source of information.

For this purpose, we denote \mathbf{e} the forecast error defined as the difference between ensemble mean forecasts and observations:

$$\mathbf{e} := \{e_1, \dots, e_T\} \quad \text{and} \quad (13)$$

$$= \{y_1 - m(x_1), \dots, y_T - m(x_T)\}, \quad (14)$$

where $m(x_t)$ and y_t are the ensemble mean and the corresponding observation at lead time $t \in \{1, \dots, T\}$, respectively. The temporal correlation of the error is described by a correlation matrix \mathbf{R}_e , defined as

$$\mathbf{R}_e = \begin{pmatrix} \rho_{e_1, e_1} & \rho_{e_1, e_2} & \dots & \rho_{e_1, e_T} \\ \rho_{e_2, e_1} & \rho_{e_2, e_2} & \dots & \rho_{e_2, e_T} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{e_T, e_1} & \rho_{e_T, e_2} & \dots & \rho_{e_T, e_T} \end{pmatrix}, \quad (15)$$

where $\rho_{e_{t_1}, e_{t_2}}$ is the correlation coefficient of the forecast error at lead times t_1 and t_2 . The empirical correlation matrix $\hat{\mathbf{R}}_e$ is estimated based on the training samples used for the univariate calibration step at the different lead times. In our setup, $\hat{\mathbf{R}}_e$ is regularly updated on a daily basis from the moving windows of 45 days defined as training datasets for the EMOS application.

Again here, we aim to construct a template [Eq. (7)] in order to establish the correlation structures within the calibrated ensemble: $\mathbf{q} := \{(q_1^1, \dots, q_1^{N_e}), \dots, (q_T^1, \dots, q_T^{N_e})\}$. In the d-ECC approach, the template is built by performing the following steps:

- 1) Apply ECC with the original ensemble forecast \mathbf{x} as the reference sample template, in order to derive a postprocessed ensemble of scenarios $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} := \{(\tilde{x}_1^1, \dots, \tilde{x}_1^{N_e}), \dots, (\tilde{x}_T^1, \dots, \tilde{x}_T^{N_e})\}. \quad (16)$$

- 2) Derive the error correction \mathbf{c}^i imposed to each scenario i ($i \in 1, \dots, N_e$) of the reference template by this postprocessing step:

$$\mathbf{c}^i := \{c_1^i, \dots, c_T^i\} \quad \text{and} \quad (17)$$

$$= \{\tilde{x}_1^i - x_1^i, \dots, \tilde{x}_T^i - x_T^i\}. \quad (18)$$

- 3) *Transformation step*: Apply a transformation to the correction \mathbf{c}^i of each scenario based on the estimate of the error autocorrelation $\hat{\mathbf{R}}_e$ and its eigendecomposition $\hat{\mathbf{R}}_e = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1}$ in order to derive the *adjusted corrections* $\tilde{\mathbf{c}}^i$:

$$\tilde{\mathbf{c}}^i = \hat{\mathbf{R}}_e^{1/2} \mathbf{c}^i \quad \text{and} \quad (19)$$

$$= \mathbf{U}\mathbf{\Lambda}^{1/2} \mathbf{U}^{-1} \mathbf{c}^i. \quad (20)$$

- 4) Derive the so-called adjusted ensemble $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} := \{(\tilde{x}_1^1, \dots, \tilde{x}_1^{N_e}), \dots, (\tilde{x}_T^1, \dots, \tilde{x}_T^{N_e})\}, \quad (21)$$

where a scenario $\tilde{\mathbf{x}}^i = \{\tilde{x}_1^i, \dots, \tilde{x}_T^i\}$ of $\tilde{\mathbf{x}}$ is defined as a combination of the original member and the adjusted error correction, namely,

$$\tilde{\mathbf{x}}^i = \mathbf{x}^i + \tilde{\mathbf{c}}^i. \quad (22)$$

- 5) Take $\tilde{\mathbf{x}}$ as the reference template in Eq. (7) so that the new empirical copula is based on the adjusted ensemble.

The d-ECC reference template $\tilde{\mathbf{x}}$ combines the raw ensemble structure and the autocorrelation of the forecast error reflected in the adjusted member corrections.

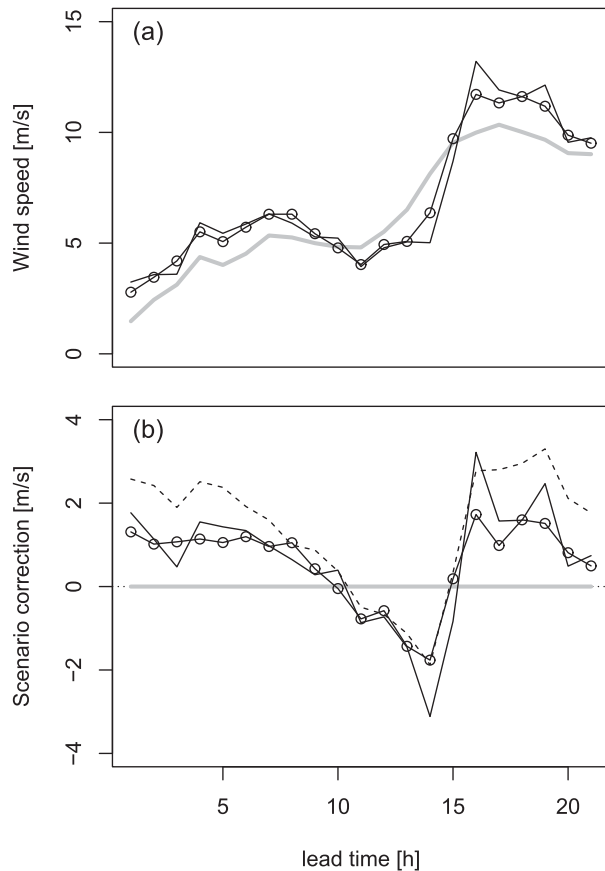


FIG. 4. Illustration of the concept of d-ECC based on the example in Fig. 3 showing (a) 1 among the 20 scenarios and (b) the correction applied to the original scenario after postprocessing. The raw ensemble forecast (here member 13) is represented in gray, the ECC scenario in black, and the d-ECC scenario in black with clear circles. The dashed line represents the scenario correction adjusted by the transformation step (see text).

The transformation of the scenario corrections in Eq. (20) adjusts their correlation structure based on the error correlation matrix $\hat{\mathbf{R}}_e$. Taking the square root of the correlation matrix [Eq. (20)] resembles a signal processing technique, which is described as a *coloring transformation* of a vector of random variables (Kessy et al. 2015).

4. Illustration and discussion of d-ECC

Focusing on a single member, the d-ECC steps are illustrated in Fig. 4. First, the correction associated with each ECC scenario with respect to the corresponding original ensemble member is computed [black line in Fig. 4b; Eq. (18)]. This scenario correction is adjusted based on the assumption of temporal autocorrelation of the error [dashed line in Fig. 4b; Eq. (20)]. This adjusted scenario correction is then superimposed onto the original ensemble forecast before the correlation structure of the adjusted ensemble is drawn again.

The new scheme reduces to the standard ECC in the case where $\text{rank}(\tilde{x}_t^i) = \text{rank}(\tilde{x}_t^i)$ for all $i \in \{1, \dots, N_e\}$ and $t \in \{1, \dots, T\}$, which means that the additional terms $\tilde{\mathbf{c}}^i$ do not have any impact on the rank structure of the ensemble. This case occurs if the following conditions are met:

- $\hat{\mathbf{R}}_e = \mathbf{I}$, where \mathbf{I} is the identity matrix, which means that there is no temporal correlation of the error in the original ensemble;
- $\mathbf{c} = \mathbf{0}$, where $\mathbf{0}$ is the null vector, which means that the calibration step does not impact the forecast, the forecast being already well calibrated; and
- $\mathbf{c} = h \times \mathbf{J}$, where h is a constant and \mathbf{J} an all-ones vector, which means that the calibration step corrects only for bias errors and the system is spread-bias free.

So the d-ECC typically takes effect if calibration corrects the spread and if this correction is correlated in time at the member level.

Additional insight can be gained by looking at the following equations. Let the observation y_t and the postprocessed ensemble members \tilde{x}_t^i be realizations of random variables \mathbf{Y} and $\tilde{\mathbf{X}}$. Consider the covariance of the forecast error denoted k and defined as

$$k_{t_1, t_2} := \mathbb{E}\{[Y_{t_1} - m(\tilde{X}_{t_1})][Y_{t_2} - m(\tilde{X}_{t_2})]\}, \quad (23)$$

where t_1 and t_2 are two lead times and $\mathbb{E}[\cdot]$ the expectation operator. It is assumed that the postprocessed ensemble mean $m(\tilde{x}_t)$ is fully bias corrected so that $\mathbb{E}[Y_t - m(\tilde{X}_t)] = 0$.

After postprocessing, the forecast scenarios and observation time series are considered as drawn from the same multivariate probability distribution, so the forecast error covariance can also be expressed as

$$k_{t_1, t_2} = \mathbb{E}\{[\tilde{X}_{t_1} - m(\tilde{X}_{t_1})][\tilde{X}_{t_2} - m(\tilde{X}_{t_2})]\} \quad (24)$$

$$= \rho_{\tilde{x}_{t_1}, \tilde{x}_{t_2}} \sigma_{\tilde{x}_{t_1}} \sigma_{\tilde{x}_{t_2}}, \quad (25)$$

where $\rho_{\tilde{x}_{t_1}, \tilde{x}_{t_2}}$ refers to the correlation between \tilde{x}_{t_1} and \tilde{x}_{t_2} , and $\sigma_{\tilde{x}_t}$ refers to the square root of the variances between the members of the calibrated ensemble $(\tilde{x}^1, \dots, \tilde{x}^{N_e})$ at lead time t . The corresponding estimators are

$$\hat{k}_{t_1, t_2} = \frac{1}{N_e - 1} \sum_{i=1}^{N_e} \{[\tilde{x}_{t_1}^i - m(\tilde{x}_{t_1})][\tilde{x}_{t_2}^i - m(\tilde{x}_{t_2})]\}, \quad (26)$$

$$\hat{\sigma}_{\tilde{x}_t} = \sqrt{\frac{1}{N_e - 1} \sum_{i=1}^{N_e} [\tilde{x}_t^i - m(\tilde{x}_t)]^2}, \quad (27)$$

and

$$\hat{\rho}_{\tilde{x}_{t_1}, \tilde{x}_{t_2}} = \frac{\hat{k}_{1,t_2}}{\hat{\sigma}_{\tilde{x}_{t_1}} \hat{\sigma}_{\tilde{x}_{t_2}}}. \quad (28)$$

From Eq. (18) recall that

$$\tilde{x}_t^i = x_t^i + c_t^i; \quad (29)$$

so, we can rewrite the expression in Eq. (25) as

$$\rho_{\tilde{x}_{t_1}, \tilde{x}_{t_2}} \sigma_{\tilde{x}_{t_1}} \sigma_{\tilde{x}_{t_2}} = \rho_{x_{t_1}, x_{t_2}} \sigma_{x_{t_1}} \sigma_{x_{t_2}} + \rho_{c_{t_1}, c_{t_2}} \sigma_{c_{t_1}} \sigma_{c_{t_2}} + \varepsilon, \quad (30)$$

where $\rho_{x_{t_1}, x_{t_2}}$ is the error autocorrelation in the original ensemble, $\rho_{c_{t_1}, c_{t_2}}$ the autocorrelation of the corrections, and σ_{x_t} and σ_{c_t} the standard deviation of the original ensemble and the standard deviation of the correction at lead time t , respectively. The term ε corresponds to the estimated covariances of x and c , and can be considered to be negligible assuming that the original forecast and the corrections are drawn from two independent random processes.

Furthermore, the stationarity assumption of d-ECC implies that the correlation $\rho_{\tilde{x}_{t_1}, \tilde{x}_{t_2}}$ can also be estimated from past error statistics:

$$\rho_{\tilde{x}_{t_1}, \tilde{x}_{t_2}} = \mathbb{E}[\hat{\rho}_{e_{t_1}, e_{t_2}}], \quad (31)$$

where the notation $\hat{\rho}_{e_{t_1}, e_{t_2}}$ refers to the elements of the estimated correlation matrix $\hat{\mathbf{R}}_{\mathbf{e}}$. The stationarity assumption takes effect in the transformation step of d-ECC [Eq. (20)], which modifies the correlation of the scenario corrections $\rho_{c_{t_1}, c_{t_2}}$ and pushes it toward the estimated correlation $\hat{\rho}_{e_{t_1}, e_{t_2}}$. In other words, the transformation affects $\rho_{c_{t_1}, c_{t_2}} \sigma_{c_{t_1}} \sigma_{c_{t_2}}$ [second term in Eq. (30)]. We expect d-ECC to have a relevant impact if $\rho_{c_{t_1}, c_{t_2}} \sigma_{c_{t_1}} \sigma_{c_{t_2}}$ dominates the sum in Eq. (30). Typically, this is the case when the spread σ_{x_t} of the original ensemble differs from the spread $\sigma_{\tilde{x}_t}$ after calibration.

To illustrate the impact of the transformation step on the correlation structure of the reference template, the scenario-generation techniques are applied to a basic synthetic dataset. For this purpose, we consider that observations and forecasts are drawn from bivariate normal distributions noted $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with $\boldsymbol{\mu}$ the mean vector and $\boldsymbol{\Sigma}$ the covariance matrix. The mean vector is set to a null vector,

$$\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

in all cases. The covariance matrix of the observation distribution is set to

$$\boldsymbol{\Sigma}_{\text{obs}} = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix},$$

so the distribution has unit variances and a correlation coefficient of 0.5 between the two dimensions. Using this setting results in target quantiles of the calibration process that correspond to the quantiles of the standard normal distribution. The covariance matrix of the forecast distribution is defined as

$$\boldsymbol{\Sigma}_{\text{fct}} = \begin{pmatrix} \alpha & \beta\alpha \\ \beta\alpha & \alpha \end{pmatrix}$$

with α a spread parameter and β a correlation parameter that allow us to simulate deficiencies in spread and correlation of the synthetic ensemble forecasts. Post-processing using ECC and d-ECC is applied considering 50 ensemble members and a sample of 1000 cases. The impact of the multivariate postprocessing schemes is illustrated by plotting the correlation coefficient between the two dimensions of the process for a range of α and β parameters (Fig. 5). The correlation coefficient of the observation is maintained as a constant (0.5) and the correlation of the raw forecasts is modified by varying the parameter β from 0.1 to 0.9. The spread parameter α takes a value of 0.5 to simulate an underdispersive ensemble, 1 a calibrated ensemble, and 1.5 an overdispersive ensemble.

The correlation structure of the forecast is not modified by applying ECC, as illustrated by the gray dashed line while the gray line shows how the transformation step affects the correlation structure of the forecast: the correlation is increased when the ensemble is underdispersed and decreased in cases of overdispersion. We find that d-ECC appears to be appropriate in the cases of ensemble forecasts with the following combination of characteristics: underdispersion combined with a lack of autocorrelation or overdispersion combined with too strong autocorrelation in the time series.

This investigation could certainly be extended considering more complex idealized studies and developing a rigorous mathematical framework. This would be welcomed as further research and would add additional evidence to the expected behavior of d-ECC. Furthermore, in the remainder of this paper, time series derived with d-ECC are compared to ECC-derived scenarios. A complementary study could aim to estimate the benefits of the dual approach with respect to purely statistical methods that only account for error characteristics estimated from historical data (Pinson et al. 2009; Möller et al. 2013).

Another important aspect of d-ECC is the estimation of the correlation matrix $\hat{\mathbf{R}}_{\mathbf{e}}$. By means of this matrix, the

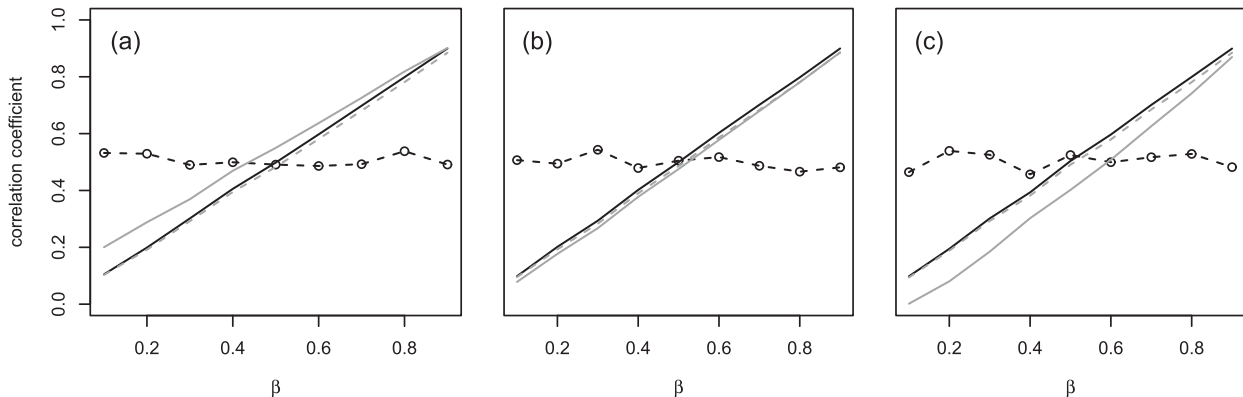


FIG. 5. Correlation coefficients between the two dimensions of the synthetic bivariate datasets as a function of the correlation parameter β with the spread parameter $\alpha =$ (a) 0.5, (b) 1, and (c) 1.5. The dashed lines with clear circles correspond to the observations, the solid black lines to the raw ensemble, the gray dashed lines to the ECC ensemble, and the gray solid lines to the d-ECC ensemble.

assumption of error autocorrelation is checked and adjusted. The matrix is estimated from the training datasets used for calibration at the different lead times. Based on the dataset described in section 2, Fig. 6 shows the lagged correlation of the forecast error derived from $\hat{\mathbf{R}}_e$. The correlation is decreasing as a function of the time lag, reaching near-zero values for lags greater than 10 h. However, for short and very short time lags, the correlation is high and stable over the rolling training datasets. In particular, focusing on a time lag of 1 h, the correlation ranges between 60% and 80%. The correlation variability shown in Fig. 6 is estimated over a 3-month period. Similar results are obtained when checking the variability of the correlation within each training dataset (not shown). The exhibited low variability indicates that the temporal correlation of the forecast error is not flow dependent. As a consequence, d-ECC can be seen as a “universal” approach that does not suffer restriction related to the forecasted weather situation.

Considering again our case study, the scenarios generated with d-ECC based on the COSMO-DE-EPS forecasts are shown in Fig. 3c. The d-ECC-derived scenarios are smoother and subjectively more realistic than the ones derived with ECC in Fig. 3b. In Fig. 4, focusing on a single scenario, it is highlighted that the difference between the original and the d-ECC time trajectories varies gradually from one time interval to the next, while abrupt transitions occur in the case of the ECC scenario, as in this example between hours 15 and 17.

Note that d-ECC does not give the same result as would a simple smoothing of the calibrated scenarios $\tilde{\mathbf{x}}$. Smoothing in time would modify the values \mathbf{q} of the calibrated ensemble and possibly diminish its reliability. Instead, d-ECC affects the time variability of the scenarios

by constructing a template [Eq. (7)] based on $\tilde{\mathbf{x}}$ [Eq. (22)] while preserving the calibrated values \mathbf{q} .

5. Verification methods

a. Multivariate scores

Verification of scenarios is first performed by assessing the multivariate aspects of the forecast by means of adequate scores. The scores are applied with a focus on

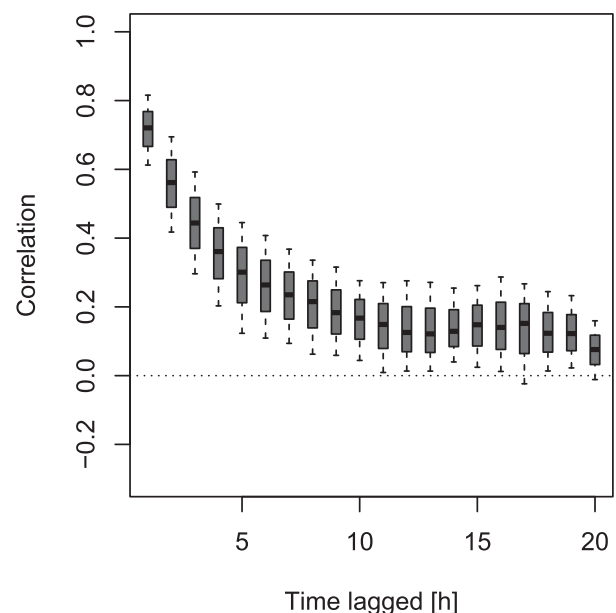


FIG. 6. Temporal lagged correlation coefficients summarizing the error correlation matrix $\hat{\mathbf{R}}_e$ used in the d-ECC approach. The box-and-whisker plots indicate the variability within the 3-month calibration period as function of lag time: the boxes cover the 25%–75% quantiles, the black line shows the 50% quantiles, and the whiskers extend to the 5%–95% quantiles.

scenarios in the form of time series. Considering an ensemble with N_e scenarios $\mathbf{x}^{(n)}$ with $n \in \{1, \dots, N_e\}$ and an observed scenario \mathbf{y} , the energy score (ES; Gneiting et al. 2008) is defined as

$$\text{ES} = \frac{1}{N_e} \sum_{n=1}^{N_e} \|\mathbf{y} - \mathbf{x}^{(n)}\| - \frac{1}{2N_e^2} \sum_{m=1}^{N_e} \sum_{p=1}^{N_e} \|\mathbf{x}^{(m)} - \mathbf{x}^{(p)}\|, \quad (32)$$

where $\|\cdot\|$ represents the Euclidean norm. ES is a generalization of the CRPS to the multivariate case.

Th eES suffers from a lack of sensitivity to misrepresentation of correlation structures (Pinson and Tastu 2013). We consider therefore additionally the p-variogram score (pVS; Scheuerer and Hamill 2015), which has better discriminative properties in this respect. Based on the geostatistical concept of variogram, pVS is defined as

$$\text{pVS} = \sum_{i \neq j} \omega_{ij} \left[|y_i - y_j|^p - \frac{1}{N_e} \sum_{n=1}^{N_e} |x_i^{(n)} - x_j^{(n)}|^p \right]^2 \quad (33)$$

with p the order of the variogram and where ω_{ij} are weights and the indices i and j indicate the i th and the j th components of the marked vectors, respectively. To focus on rapid changes in wind speed, the weights ω_{ij} are chosen proportional to the inverse square distance in time such that

$$\omega_{ij} = \frac{1}{(i-j)^2}, \quad i \neq j, \quad (34)$$

since i and j are here forecast lead-time indices.

b. Multivariate rank histograms

The multivariate aspect of the forecast is in a second step assessed by means of rank histograms applied to multidimensional fields (Thorarinsdottir et al. 2016). Two variants of the multivariate rank histogram are applied: the averaged rank histogram (ARH) and the band depth rank histogram (BDRH). The difference between the two approaches lies in the way they distinguish pre-ranks from multivariate forecasts. ARH considers the averaged rank over the multivariate aspect while BDRH assesses the centrality of the observation within the ensemble based on the concept of functional band depth.

The interpretation of ARH is the same as the interpretation of a univariate rank histogram: U-shaped, \cap -shaped and flat rank histograms are interpreted as underdispersion, overdispersion, and calibration of the underlying ensemble forecasts, respectively. The interpretation of BDRH is different: a \cup shape is associated

with a lack of correlation, an \cap shape to a too high correlation in the ensemble, a skewed rank histogram to bias or dispersion errors and a flat rank histogram to calibrated forecasts.

c. Product-oriented verification

In addition to multivariate verification of time series scenarios, the forecasts are assessed within a product-oriented framework. This type of scenario verification follows the spirit of the event-oriented verification framework proposed by Pinson and Girard (2012). Probabilistic forecasts that require time trajectories are provided and assessed by means of well-established univariate probabilistic scores.

Two types of products derived from forecasted scenarios are examined here. The first one is defined as the mean wind speed over a day (here, a day is limited to the 21-h forecast horizon). The second product is defined as the maximal upward wind ramp over a day, a wind ramp being defined as the difference between two consecutive forecast intervals. For both products, 20 forecasts are derived from the 20 scenarios at each station and each verification day.

The performances of the ensemble forecasts for the two types of products are evaluated by means of the CRPS. The CRPS is the generalization of the mean absolute error to predictive distributions (Gneiting et al. 2008), and can be seen as the integral of the Brier score (BS; Brier 1950) over all thresholds or the integral of the quantile score (QS; Koenker and Bassett 1978) over all probability levels. Considering an ensemble forecast, the CRPS can be calculated as a weighted sum of QS applied to the sorted ensemble members (Bröcker 2012). For more insight into forecast performance in terms of attributes, the CRPS is decomposed following the same approach (Ben Bouallègue 2015): the CRPS reliability and resolution components are calculated as weighted sums of the reliability and resolution components of the QS at the probability levels defined by the ensemble size [see Eq. (3)], respectively. Formally, we write

$$\text{CRPS}_{\text{reliability}} = \frac{2}{N_e} \sum_{n=1}^{N_e} \text{QS}_{\text{reliability}}^{(\tau_n)} \quad \text{and} \quad (35)$$

$$\text{CRPS}_{\text{resolution}} = \frac{2}{N_e} \sum_{n=1}^{N_e} \text{QS}_{\text{resolution}}^{(\tau_n)}, \quad (36)$$

where $\text{QS}_{\text{reliability}}^{(\tau_n)}$ and $\text{QS}_{\text{resolution}}^{(\tau_n)}$ are the reliability and resolution components, respectively, of the QS applied to the quantile forecasts at probability level τ_n . The QS decomposition is performed following Bentzien and

Friederichs (2014). The $\text{CRPS}_{\text{reliability}}$ is negatively oriented (the lower the better) while the $\text{CRPS}_{\text{resolution}}$ is positively oriented (the higher the better).

d. Bootstrapping

The statistical significance of the results is tested by applying a block-bootstrap approach. Bootstrapping is a resampling technique that provides an estimation of the statistical consistency and is commonly applied to meteorological datasets (Efron and Tibshirani 1986).

A block-bootstrap approach is applied in the following, which consists of defining a block as a single day during the verification period (Hamill 1999). Each day is considered to be a separate block of fully independent data. The verification process is repeated 500 times by using each time a random sample with replacement of the 92 verification days (March–May 2013). The derived score distributions illustrate consequently the variability of the performance measures over the verification period and not between locations. Boxplots are used to represent the distributions of the performance measures, where the quantile of the distributions at probability levels of 5%, 25%, 50%, 75%, and 95% are highlighted.

6. Results and discussion

Before applying the verification methods introduced in the previous section, we propose to explore statistically the COSMO-DE-EPS time series variability by means of a spectral analysis, an analysis of the time series in the frequency domain. Such an analysis is useful in order to describe the statistical properties of the scenarios but also has direct implications for user's applications (see below; Vincent et al. 2010). A Fourier transformation is applied to each forecasted and observed scenario and the contributions of the oscillations at various frequencies to the scenario variance are examined (Wilks 2006). In Fig. 7, the mean amplitude of the forecast and observation time series over all stations and verification days is plotted as a function of their frequency components.

As has already been suggested by the case study, this analysis confirms that the ECC considerably increases the variability of the time trajectories with respect to the original ensemble, in particular at high frequencies. The ECC scenario fluctuations are also much larger than the observed ones. Indeed, the amplitude is on average about 2 times larger at high frequencies in ECC time series than in the observed results, which explains the visual impression that ECC scenarios are unrealistic. Conversely, scenarios derived with the new copula approach do not exhibit such features. While the original ensemble shows a deficit of variability with

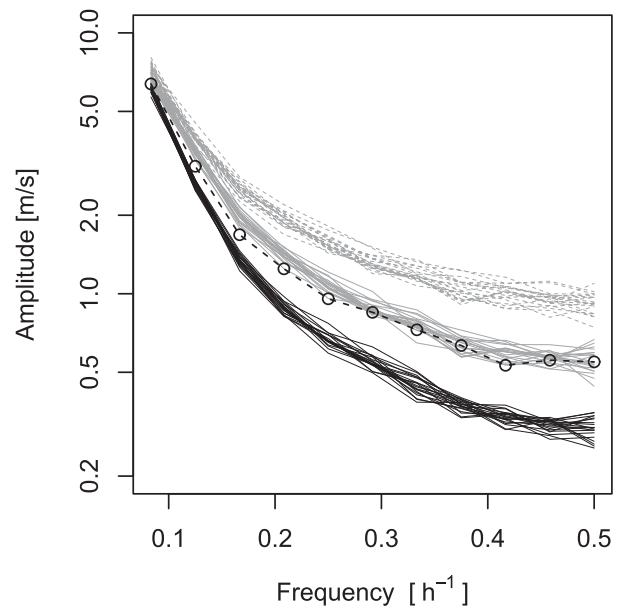


FIG. 7. Spectral analysis of the scenarios from the raw ensemble (black lines) of the scenarios derived with ECC (dashed gray lines) and with d-ECC (gray lines). Each line corresponds to 1 scenario among the 20. The spectrum of the observed time series is represented by the dashed line with clear circles.

respect to the observations, the d-ECC approach allows for improving this aspect of the forecast. This first result, showing that d-ECC scenarios have a mean spectrum similar to that of the observations, is complemented with an objective assessment of the forecasted scenarios based on probabilistic verification measures.

Figure 8 shows the performance of the forecasted time trajectories by means of multivariate scores. The postprocessed scenarios perform significantly better than the raw members in terms of ES (Fig. 8a). In terms of pVS, the d-ECC scenarios are better than the ECC ones and significantly better than the raw ones when $p = 0.5$ (Fig. 8b). For higher orders of the variogram (here $p = 1$; Fig. 8c), the forecast improvement after postprocessing is still clear when using d-ECC while the ECC results are slightly worse than those of the original forecasts.

Figure 9 depicts the results in terms of multivariate rank histograms for ARH (top panel) and BDRH (bottom panel). The raw ensemble shows clear reliability deficiencies (Figs. 9a,d), which motivated the use of postprocessing techniques. Forecasts derived with ECC continue to show underdispersiveness but also too little correlation (Figs. 9b,e) while forecasts derived with d-ECC are better calibrated according to the rank histograms in Figs. 9c,f. Indeed, both plots indicate good reliability among the d-ECC-derived scenarios.

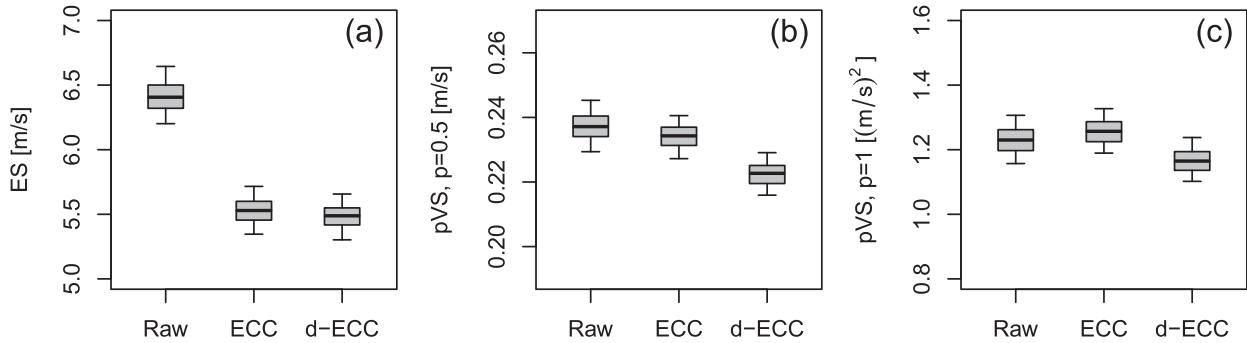


FIG. 8. Multivariate scores of time series: (a) energy score and (b),(c) p-variogram scores for $p = 0.5$ and 1 , respectively, in the form of box plots drawn from the application of a 500-block bootstrapping for the raw, ECC, and d-ECC. The box-and-whisker plots indicate the 25%–75% and 5%–95% confidence intervals, respectively.

Figure 10 focuses on two products drawn from the time series forecasts: the daily mean wind speed (top panel) and the daily maximal upward ramp (bottom panel). The performance are assessed in terms of CRPS, CRPS reliability, and CRPS resolution, from left to right, respectively. Looking at the results in terms of CRPS, we note the high degree of similarity between Figs. 10a and 10d and Figs. 8a and 8c. As for the ES, postprocessing significantly improves the forecasts of the daily mean product. As for pVS with $p = 1$, d-ECC improves the ramp product with respect to the original while ECC does not generate improved products. The CRPS decomposition allows us to provide detail related to the origin of these performance improvements. We see in Figs. 10b,e that the CRPS results are mainly explained by the impact of the postprocessing on the CRPS reliability components. However, focusing on the results in terms of CRPS resolution in Figs. 10c,f, we note that the resolution of the original and d-ECC

products are comparable while ECC deteriorates the resolution of the ramp product with respect to the original.

These verification results are interpreted as follows. Calibration corrects for the mean of the ensemble forecast and this is reflected, after the derivation of scenarios, by an improvement in the ES and daily mean product skill. Calibration also corrects for spread deficiencies increasing the variability of the ensemble forecasts. This increase in spread associated with the preservation of the rank structure of the original ensemble, as is the case with the ECC approach, enlarges indiscriminately the temporal variability of the forecasts and leads to a slight deterioration of the pVS and ramp product results.

The d-ECC approach provides scenarios with a temporal variability comparable to that of the observations. In that case, the benefit of the calibration step in terms of reliability (at single forecast lead times) persists at the multivariate level (looking at time trajectories) after the reconstruction of scenarios with d-ECC. The multivariate reliability, or the

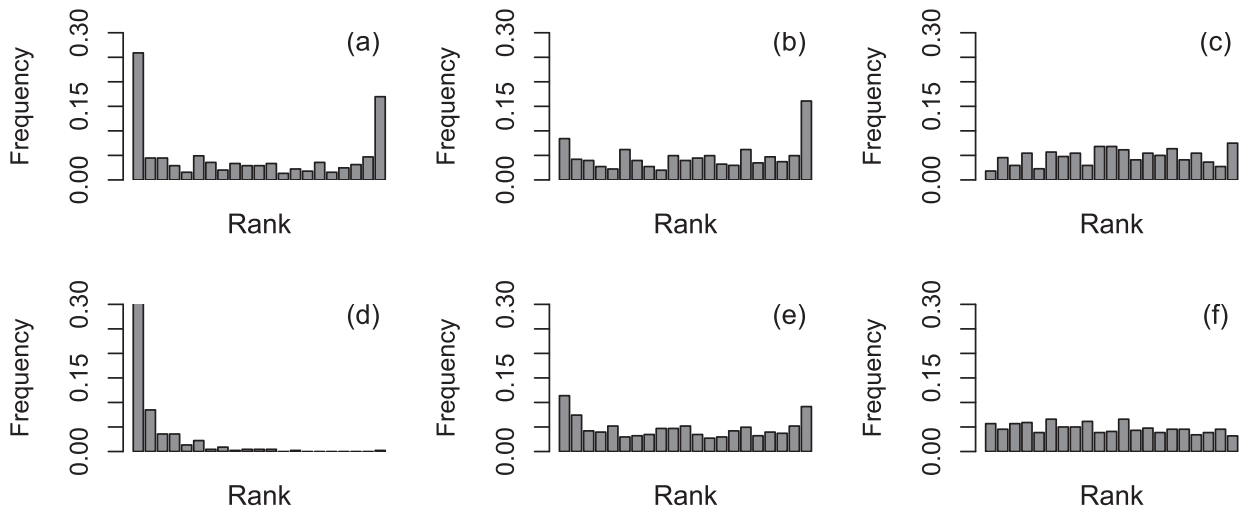


FIG. 9. Multivariate rank histograms: (a)–(c) average rank and (d)–(f) band depth rank for (a),(d) time series from the raw ensemble and derived with (b),(e) ECC and (c),(f) d-ECC.

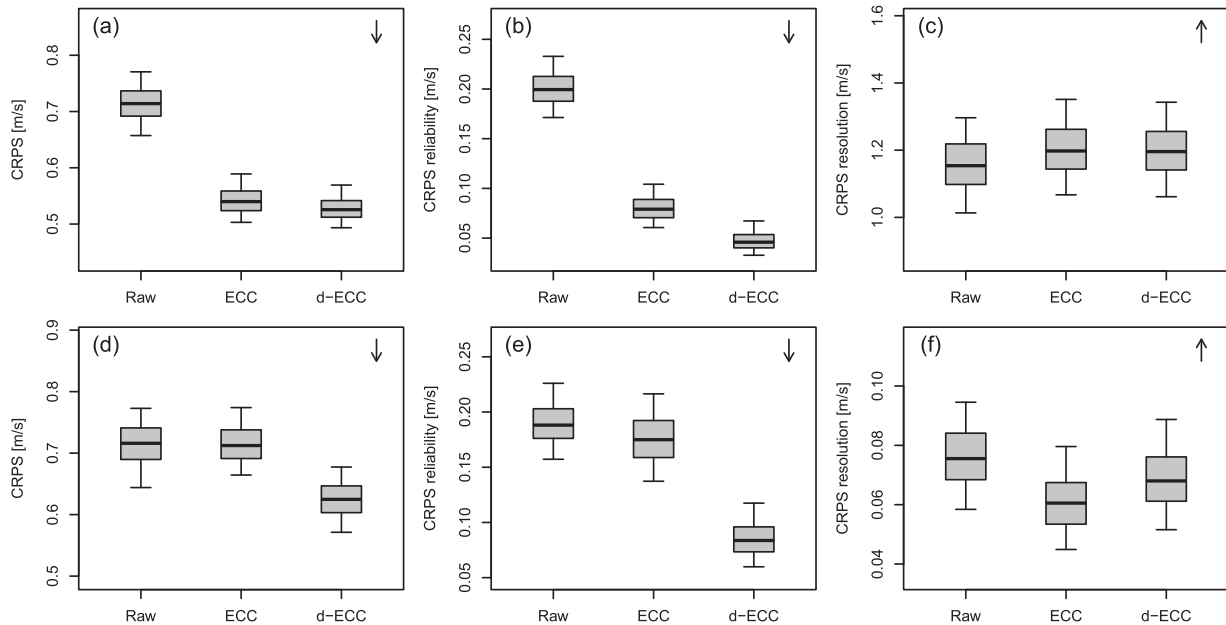


FIG. 10. Product-oriented verification of scenarios: (a)–(c) daily means at stations, (d)–(f) maximal upward ramps within a day at a station. Results are shown in terms of (a),(d) CRPS, (b),(e) CRPS reliability component, and (c),(f) CRPS resolution component. The box plots indicate confidence intervals estimated with block bootstrapping. The arrows in the right corners indicate whether the performance measure is positively or negatively oriented. The box-and-whisker plots indicate the 25%–75% and 5%–95% confidence intervals, respectively.

reliability of derived products, is significantly improved after postprocessing, though it is not perfect for specific derived products. Moreover, d-ECC scenarios perform as well as the original ensemble forecast in terms of resolution. So, unlike ECC, d-ECC is able to generate reliable scenarios with a level of resolution that does not deteriorate with respect to the original ensemble forecasts.

7. Conclusions and outlook

A new empirical copula approach is proposed for the postprocessing of calibrated ensemble forecasts. The so-called dual-ensemble copula coupling approach is introduced with a focus on temporal structures of wind forecasts. The new scheme includes a temporal component in the ECC approach accounting for the error autocorrelation of the ensemble members. The estimation of the correlation structure in the error based on past data allows for adjusting the dependence structure in the original ensemble.

Based on COSMO-DE-EPS forecasts, the scenarios derived with d-ECC prove to be qualitatively realistic and quantitatively of superior quality. Postprocessing of wind speed combining EMOS and d-ECC improves the forecasts in many aspects. In comparison to ECC, d-ECC drastically improves the quality of the derived scenarios. Applications that require temporal trajectories will fully benefit from the new approach in that case.

As for any postprocessing technique, the benefit of the new copula approach can be weakened by improving the representation of the forecast uncertainty with more efficient member generation techniques and/or by improving the calibration procedure correcting for conditional biases. Meanwhile, with its low additional complexity and computational costs, d-ECC can be considered to be a valuable alternative to the standard ECC for the generation of consistent scenarios from COSMO-DE-EPS.

Though only the temporal aspects have been investigated in this study, the dual-ensemble copula approach could be generalized to any multivariate setting. Further research is however required for the application of d-ECC at scales that are unresolved by the observations. For example, geostatistical tools could be applied for the description of the autocorrelation error structure at the model grid level. Moreover, the mathematical interpretation of the d-ECC scheme developed here would benefit from further theoretical investigation.

Acknowledgments. This work has been done within the framework of the EWeLiNE project (Erstellung innovativer Wetter- und Leistungsprognosemodelle für die Netzintegration wetterabhängiger Energieträger) funded by the German Federal Ministry for Economic Affairs and Energy. The authors acknowledge the

Department of Wind Energy of the Technical University of Denmark (DTU), the German Wind Energy Institute (DEWI GmbH), DNV GL, the Meteorological Institute (MI) of University of Hamburg, and the Karlsruhe Institute of Technology (KIT) for providing wind measurements at stations Risoe, *FINO1* and *FINO3*, *FINO2*, Hamburg and Lindenberg, and Karlsruhe, respectively. The authors are also grateful to Tilmann Gneiting and two anonymous reviewers for helpful and accurate comments on a previous version of this manuscript.

REFERENCES

- Ben Bouallègue, Z., 2013: Calibrated short-range ensemble precipitation forecasts using extended logistic regression with interaction terms. *Wea. Forecasting*, **28**, 515–524, doi:[10.1175/WAF-D-12-00062.1](https://doi.org/10.1175/WAF-D-12-00062.1).
- , 2015: Assessment and added value estimation of an ensemble approach with a focus on global radiation forecasts. *Mausam*, **66**, 541–550.
- Bentzien, S., and P. Friederichs, 2014: Decomposition and graphical portrayal of the quantile score. *Quart. J. Roy. Meteor. Soc.*, **140**, 1924–1934, doi:[10.1002/qj.2284](https://doi.org/10.1002/qj.2284).
- Bremnes, J. B., 2004: Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, **7**, 47–54, doi:[10.1002/we.107](https://doi.org/10.1002/we.107).
- Brier, G. W., 1950: Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, **78**, 1–3, doi:[10.1175/1520-0493\(1950\)078<0001:VOFEIT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2).
- Bröcker, J., 2012: Evaluating raw ensembles with the continuous ranked probability score. *Quart. J. Roy. Meteor. Soc.*, **138**, 1611–1617, doi:[10.1002/qj.1891](https://doi.org/10.1002/qj.1891).
- Clark, M., S. Gangopadhyay, L. Hay, B. Rajagopalan, and R. Wilby, 2004: The Schaake shuffle: A method for reconstructing space–time variability in forecasted precipitation and temperature fields. *J. Hydrometeorol.*, **5**, 243–262, doi:[10.1175/1525-7541\(2004\)005<0243:TSSAMF>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0243:TSSAMF>2.0.CO;2).
- Efron, B., and R. Tibshirani, 1986: Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Stat. Sci.*, **1**, 54–75, doi:[10.1214/ss/1177013815](https://doi.org/10.1214/ss/1177013815).
- Feldmann, K., M. Scheuerer, and T. Thorarinsdottir, 2015: Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Wea. Rev.*, **143**, 955–971, doi:[10.1175/MWR-D-14-00210.1](https://doi.org/10.1175/MWR-D-14-00210.1).
- Flowerdew, J., 2014: Calibrating ensemble reliability whilst preserving spatial structure. *Tellus*, **66A**, 22662, doi:[10.3402/tellusa.v66.22662](https://doi.org/10.3402/tellusa.v66.22662).
- Gebhardt, C., S. E. Theis, M. Paulat, and Z. Ben Bouallègue, 2011: Uncertainties in COSMO-DE precipitation forecasts introduced by model perturbations and variation of lateral boundaries. *Atmos. Res.*, **100**, 168–177, doi:[10.1016/j.atmosres.2010.12.008](https://doi.org/10.1016/j.atmosres.2010.12.008).
- Gneiting, T., F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration, and sharpness. *J. Roy. Stat. Soc.*, **69B**, 243–268, doi:[10.1111/j.1467-9868.2007.00587.x](https://doi.org/10.1111/j.1467-9868.2007.00587.x).
- , L. Stanberry, E. Grimit, L. Held, and N. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds. *Test*, **17**, 211–235, doi:[10.1007/s11749-008-0114-x](https://doi.org/10.1007/s11749-008-0114-x).
- Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting*, **14**, 155–167, doi:[10.1175/1520-0434\(1999\)014<0155:HTFENP>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0155:HTFENP>2.0.CO;2).
- Kessy, A., A. Lewin, and K. Strimmer, 2015: Optimal whitening and decorrelation. *arXiv.org*, 14 pp. [Available online at <https://arxiv.org/abs/1512.00809>.]
- Keune, J., C. Ohlwein, and A. Hense, 2014: Multivariate probabilistic analysis and predictability of medium-range ensemble weather forecasts. *Mon. Wea. Rev.*, **142**, 4074–4090, doi:[10.1175/MWR-D-14-00015.1](https://doi.org/10.1175/MWR-D-14-00015.1).
- Koenker, R., and G. Bassett, 1978: Regression quantiles. *Econometrica*, **46**, 33–50, doi:[10.2307/1913643](https://doi.org/10.2307/1913643).
- Krzysztofowicz, R., 1983: Why should a forecaster and a decision maker use Bayes theorem. *Water Resour. Res.*, **19**, 327–336, doi:[10.1029/WR019i002p00327](https://doi.org/10.1029/WR019i002p00327).
- Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. *J. Comput. Phys.*, **227**, 3515–3539, doi:[10.1016/j.jcp.2007.02.014](https://doi.org/10.1016/j.jcp.2007.02.014).
- Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. *Manage. Sci.*, **22**, 1087–1096, doi:[10.1287/mnsc.22.10.1087](https://doi.org/10.1287/mnsc.22.10.1087).
- Möller, A., A. Lenkoski, and T. L. Thorarinsdottir, 2013: Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. *Quart. J. Roy. Meteor. Soc.*, **139**, 982–991, doi:[10.1002/qj.2009](https://doi.org/10.1002/qj.2009).
- Peralta, C., Z. Ben Bouallègue, S. E. Theis, and C. Gebhardt, 2012: Accounting for initial condition uncertainties in COSMO-DE-EPS. *J. Geophys. Res.*, **117**, D07108, doi:[10.1029/2011JD016581](https://doi.org/10.1029/2011JD016581).
- Pinson, P., 2012: Adaptive calibration of (u,v)-wind ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1273–1284, doi:[10.1002/qj.1873](https://doi.org/10.1002/qj.1873).
- , 2013: Wind energy: Forecasting challenges for its operational management. *Stat. Sci.*, **28**, 564–585, doi:[10.1214/13-STS445](https://doi.org/10.1214/13-STS445).
- , and R. Girard, 2012: Evaluating the quality of scenarios of short-term wind power generation. *Appl. Energy*, **96**, 12–20, doi:[10.1016/j.apenergy.2011.11.004](https://doi.org/10.1016/j.apenergy.2011.11.004).
- , and J. Tastu, 2013: Discrimination ability of the energy score. DTU Tech. Rep., Technical University of Denmark, 16 pp. [Available online at http://orbit.dtu.dk/files/56966842/tr13_15_Pinson_Tastu.pdf.]
- , G. Papaefthymiou, B. Klockl, H. Nielsen, and H. Madsen, 2009: From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, **12**, 51–62, doi:[10.1002/we.284](https://doi.org/10.1002/we.284).
- Schefzik, R., T. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, **28**, 616–640, doi:[10.1214/13-STS443](https://doi.org/10.1214/13-STS443).
- Scheuerer, M., and T. M. Hamill, 2015: Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Wea. Rev.*, **143**, 1321–1334, doi:[10.1175/MWR-D-14-00269.1](https://doi.org/10.1175/MWR-D-14-00269.1).
- Schölzel, C., and A. Hense, 2011: Probabilistic assessment of regional climate change in southwest Germany by ensemble dressing. *Climate Dyn.*, **36**, 2003–2014, doi:[10.1007/s00382-010-0815-1](https://doi.org/10.1007/s00382-010-0815-1).
- Schuhen, N., T. Thorarinsdottir, and T. Gneiting, 2012: Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, **140**, 3204–3219, doi:[10.1175/MWR-D-12-00028.1](https://doi.org/10.1175/MWR-D-12-00028.1).
- Sklar, M., 1959: Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, **8**, 229–231.
- Sloughter, J., T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model

- averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35, doi:[10.1198/jasa.2009.ap08615](https://doi.org/10.1198/jasa.2009.ap08615).
- Thorarindottir, T., and T. Gneiting, 2010: Probabilistic forecasts of wind speed: Ensemble model output statistics by using heteroscedastic censored regression. *J. Roy. Stat. Soc.*, **173A**, 371–388, doi:[10.1111/j.1467-985X.2009.00616.x](https://doi.org/10.1111/j.1467-985X.2009.00616.x).
- , M. Scheuerer, and C. Heinz, 2016: Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *J. Comput. Graph. Stat.*, **25**, 105–122, doi:[10.1080/10618600.2014.977447](https://doi.org/10.1080/10618600.2014.977447).
- Vincent, C., G. Giebel, P. Pinson, and H. Madsen, 2010: Resolving nonstationary spectral information in wind speed time series using the Hilbert–Huang transform. *J. Appl. Meteor. Climatol.*, **49**, 253–267, doi:[10.1175/2009JAMC2058.1](https://doi.org/10.1175/2009JAMC2058.1).
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. Academic Press, 627 pp.
- , 2015: Multivariate ensemble model output statistics using empirical copulas. *Quart. J. Roy. Meteor. Soc.*, **141**, 945–952, doi:[10.1002/qj.2414](https://doi.org/10.1002/qj.2414).